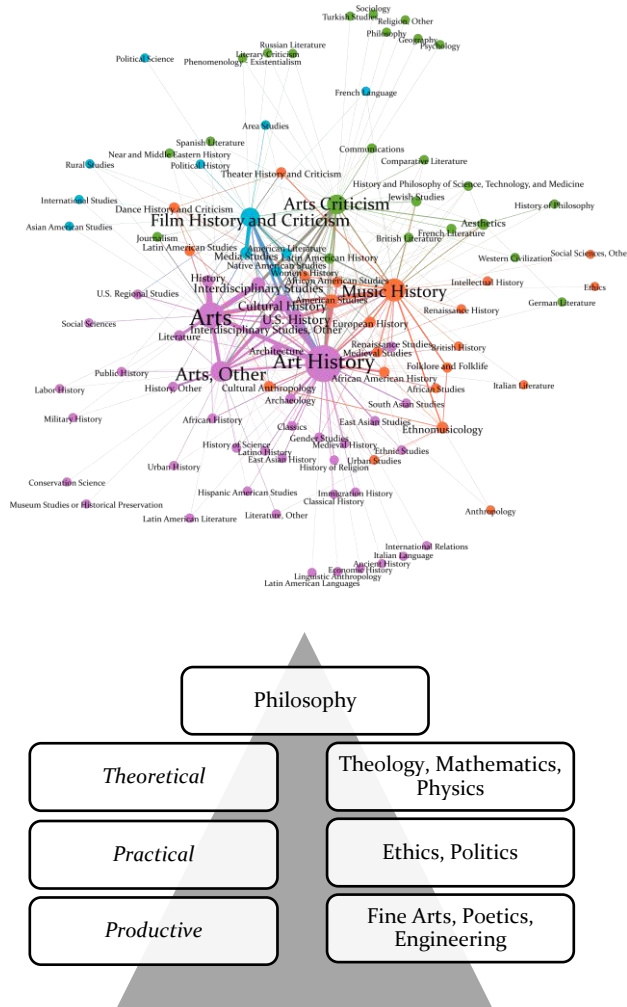
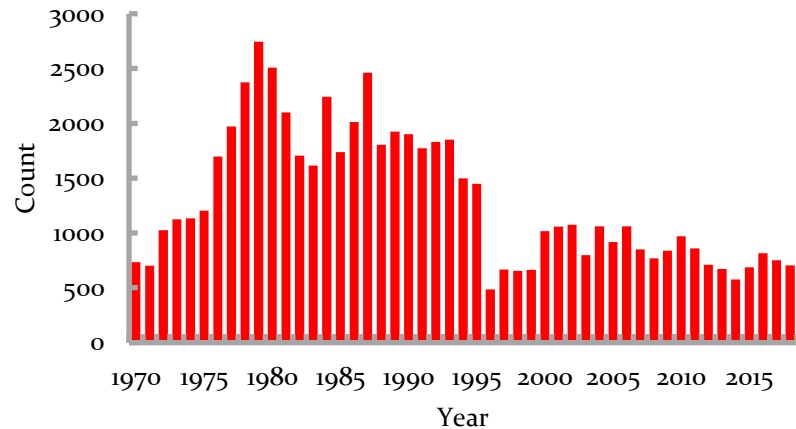


VPA 5300: The Digital Humanities in the Arts



Week 2 – 02

Tutorial: Scraping & APIs



Digital Reception (Auto)Biographies

Project Pipeline

- 1) Select (shared) media platform.
- 2) Download/extract data from platform.
- 3) Explore data set.
- 4) Pose research question(s).
- 5) Scrape/query data OR annotate.
- 6) Analyze data.
- 7) Create presentation.

Digital Media Platforms

Streaming Services

Video

- Netflix
- Hulu
- HBO Max
- Amazon Prime

Audio

- Spotify
- Pandora
- Apple Music
- YouTube Music

Social Networking Services

- Facebook
- Twitter
- Instagram
- Pinterest

Media Players

- iTunes

Example – Gender Bias on Netflix

Problems

- ~30% of lead actors in the sample were women.
- ~20% of writers in the sample were women.
- ~7% of directors in the sample were women.

– Geena Davis Institute on Gender in Media

Research Question

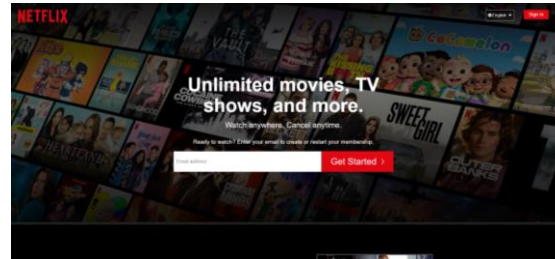
Do my viewing habits on Netflix reflect these inequities?

<https://seejane.org/research-informs-empowers/>

Data sets

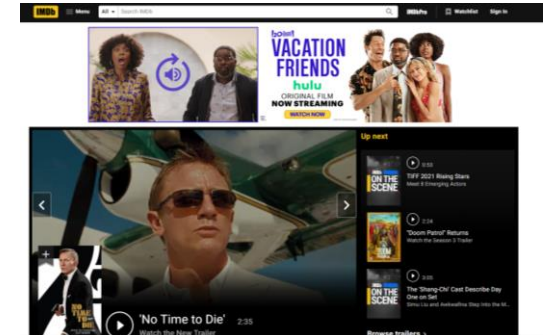
Netflix

- Viewing History
 - Title
 - Viewing Date



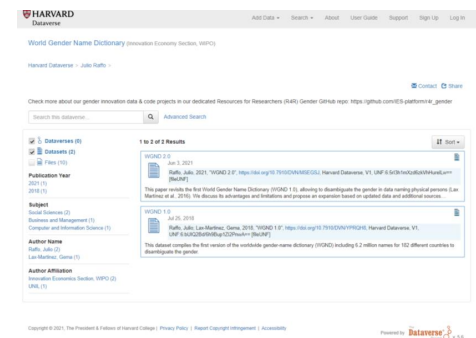
Open Movie Database (OMDb)

- Title information
 - Directors
 - Writers
 - Actors



World Gender Name Dictionary

- List of Names
 - Gender
 - Country



Web Scraping

- the extraction of data from a website.

Internet Movie Database (IMDb)

- Scrape data pertaining to the names of the cast, directors, and writers for each title in my Netflix Viewing History.
- `tutorial_scraping.m`

https://www.imdb.com/title/tto796366/?ref=fn_al_tt_2

Web Scraping

Ethics

IMDB Conditions of Use

be affiliated with, connected to, or sponsored by IMDb.

License and Site Access

Subject to your compliance with these Conditions of Use and your payment of any applicable fees, IMDb or its content providers grants you a limited, non-exclusive, non-transferable, non-sublicenseable license to access and make personal and non-commercial use of the IMDb Services, including digital content available through the IMDb Services, and not to download (other than page caching) or modify this site, or any portion of it, except with express written consent of IMDb. Additional license terms may be found in the Terms. The IMDb Services or any portion of such services may not be reproduced, duplicated, copied, sold, resold, visited, or otherwise exploited for any commercial purpose without express written consent of IMDb. This license does not include any resale or commercial use of any IMDb Service or its contents or any derivative use of this site or its contents. All licenses are non-exclusive and all rights not expressly granted to you in these Conditions of Use or any applicable Terms are reserved and retained by IMDb or its licensors, suppliers, publishers, rightsholders, or other content providers. You will use all IMDb Services in compliance with all applicable laws.

Robots and Screen Scraping: You may not use data mining, robots, screen scraping, or similar data gathering and extraction tools on this site, except with our express written consent as noted below.

Framing: You may not frame or utilize framing techniques to enclose any trademark, logo, or other proprietary information (including images,

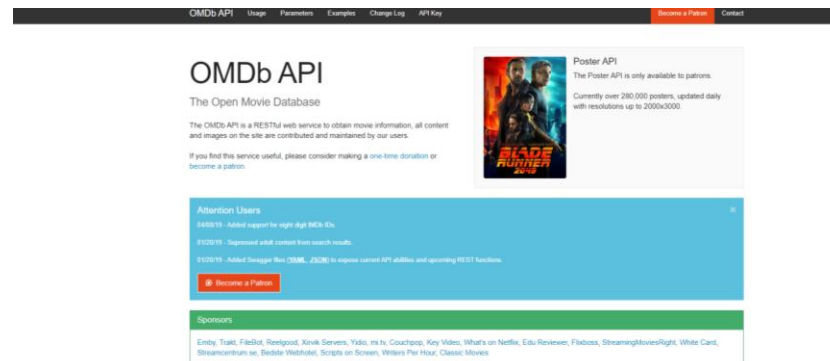
<https://www.imdb.com/conditions>

Application Programming Interfaces (APIs)

- a set of commands that can be used to retrieve the data stored in a company's database(s).

The Open Movie Database (OMDb) API

- Query data pertaining to the names of the cast, directors, and writers for each title in my Netflix Viewing History.
- `tutorial_scraping.m`

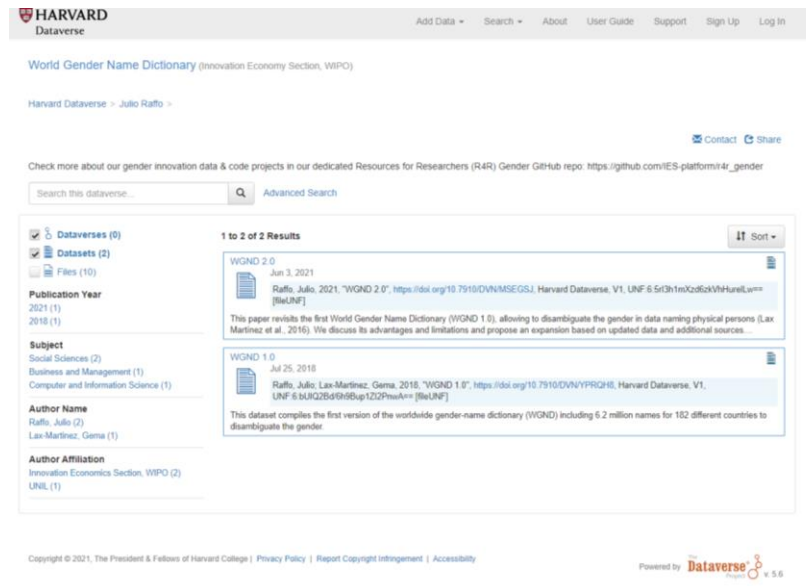


<https://www.omdbapi.com/>

Gender Identification

World Gender Name Dictionary (WGND)

- Match first names of lead actors, writers, and directors to gender identification database.
- `tutorial_scraping.m`



<https://dataverse.harvard.edu/dataverse/WGND>

Findings

Actors

- 40% of lead actors in the sample were women (30%).

Writers

- 29% of lead writers in the sample were women (20%).

Directors

- 21% of lead directors in the sample were women (7%).

Other Questions

- How does genre interact with gender representation?
- How does the presence of multiple spoken languages in a film interact with gender representation?

Next Class (Tuesday, 9/7)

- Empiricism and Epistemology

Homework:

- **READING:** Huron, “The New Empiricism...” (37 pages)

Advice: Read it in two sittings.